

Communication: An efficient stochastic algorithm for the perturbative density matrix renormalization group in large active spaces

Sheng Guo, Zhendong Li, and Garnet Kin-Lic Chan

Citation: *The Journal of Chemical Physics* **148**, 221104 (2018); doi: 10.1063/1.5031140

View online: <https://doi.org/10.1063/1.5031140>

View Table of Contents: <http://aip.scitation.org/toc/jcp/148/22>

Published by the *American Institute of Physics*

PHYSICS TODAY

WHITEPAPERS

ADVANCED LIGHT CURE ADHESIVES

Take a closer look at what these environmentally friendly adhesive systems can do

READ NOW

PRESENTED BY
MASTERBOND
ADHESIVES | SEALANTS | COATINGS

Communication: An efficient stochastic algorithm for the perturbative density matrix renormalization group in large active spaces

Sheng Guo, Zhendong Li, and Garnet Kin-Lic Chan^{a)}

Division of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, California 91125, USA

(Received 28 March 2018; accepted 31 May 2018; published online 14 June 2018)

We present an efficient stochastic algorithm for the recently introduced perturbative density matrix renormalization group method for large active spaces. The stochastic implementation bypasses the computational bottleneck involved in solving the first order equation in the earlier deterministic algorithm. We demonstrate the efficiency and accuracy of the algorithm on the C_2 and Cr_2 molecular benchmark systems. *Published by AIP Publishing.* <https://doi.org/10.1063/1.5031140>

Molecular electronic structure with a mix of static correlation and dynamic correlation remains a challenging aspect of quantum chemistry. It is often treated within a complete active space (CAS) approach for the static correlation plus a second-order perturbative treatment of the dynamic correlation, such as with CASPT2^{1,2} (complete-active-space second-order perturbation theory) and NEVPT2^{3–5} (N -electron valence state perturbation theory). However, in many cases, the second-order treatment of correlations outside of the active space is not quantitatively accurate, and it becomes necessary to enlarge the active space with orbitals of intermediate correlation strength. For instance, in $3d$ transition metal complexes, the virtual $4d$, semi-core ($3s$ and $3p$), or valence ligand orbitals often need to be included in the active space for good accuracy.^{6–8} The resulting very large number of active orbitals with a mix of different correlation character renders the density matrix renormalization group (DMRG),^{9–19} whose strength lies in treating large active spaces where all the orbitals are strongly interacting, relatively inefficient, and the subsequent dynamic correlation treatment also becomes very difficult.^{7,8,20–25}

Recently, we developed a perturbative DMRG (p-DMRG)²⁶ algorithm to efficiently target near-exact energies in large orbital spaces where there is a mix of strongly and intermediately correlated orbitals. We showed that the p-DMRG provides, in practice, benchmark quality energies as accurate as those obtained by variational DMRG calculations, but with a significantly reduced computational cost. The essential idea is to treat the strong correlation with a variational DMRG calculation with small bond dimension M_0 (which can be systematically increased) within the full orbital space, and the residual correlation by second-order perturbation theory (PT2) within the same orbital space. The total energy obtained by p-DMRG converges very quickly with M_0 , which can thus be chosen significantly smaller than the bond dimension M in the fully variational calculation. The basic idea of p-DMRG is similar in spirit to that of selected configuration

interaction (SCI) plus perturbation theory,^{27–35} and in particular, the recently developed Heat-bath CI+PT (HCI+PT),^{33,34} where the static correlation is treated by configuration interaction within a selected set of important determinants defined by a threshold, which can be tuned to systematically increase the accuracy much like M_0 in p-DMRG, with the residual dynamic correlation in the same orbital space treated by standard PT2.

In this communication, we describe an efficient stochastic algorithm that further speeds up the p-DMRG algorithm by avoiding having to solve the first order equation deterministically. Such a stochastic extension to HCI+PT has already been described.³⁴ However, in the selected CI setting, the stochastic extension is very natural due to a clear separation of the variational and first order determinant spaces. Since in p-DMRG, the zeroth order wavefunction is a matrix product state (MPS),¹⁹ the stochastic extension is less straightforward and this communication is thus concerned primarily with its correct and efficient formulation.

To begin with, we assume that a zeroth order wavefunction $|\Psi^{(0)}\rangle$, expressed as an MPS with a small bond dimension M_0 , has been optimized by the standard variational DMRG algorithm. Given a partitioning of the Hamiltonian, $\hat{H} = \hat{H}_0 + \hat{V}$ with $\hat{H}_0|\Psi^{(0)}\rangle = E_0|\Psi^{(0)}\rangle$, the first order wavefunction $|\Psi^{(1)}\rangle$ is defined by the first order equation,

$$Q(\hat{H}_0 - E_0)Q|\Psi^{(1)}\rangle = -Q\hat{V}|\Psi^{(0)}\rangle, \quad (1)$$

where $Q = 1 - P$ and $P = |\Psi^{(0)}\rangle\langle\Psi^{(0)}|$ are projectors. Once $|\Psi^{(1)}\rangle$ is obtained from Eq. (1), which in our implementation is achieved by minimizing the Hylleraas functional using the MPSPT algorithm,²⁰ the second order energy $E_2 = \langle\Psi^{(1)}|Q\hat{V}|\Psi^{(0)}\rangle$ can be computed. For very large active spaces with 50–100 orbitals, the required bond dimension M_1 for expanding $|\Psi^{(1)}\rangle$, which typically scales like $O(K^2M_0)$ (K is the number of spatial orbitals), can be quite large ($\gg 10^4$). This creates a significant cost both in computation and in storage. In the previous deterministic p-DMRG algorithm,²⁶ we used a sum of MPS representation $|\Psi^{(1)}\rangle = \sum_i |\Psi_i^{(1)}\rangle$ as well as extrapolation, which helped alleviate the computational

^{a)}Electronic mail: gkc1000@gmail.com

cost. By contrast, a stochastic algorithm can be expected to essentially eliminate this bottleneck, at the cost of introducing statistical errors.

To develop a stochastic variant of p-DMRG, we first rewrite E_2 as

$$E_2 = -\langle \Psi^{(0)} | \hat{V} Q [Q(\hat{H}_0 - E_0)Q]^{-1} Q \hat{V} | \Psi^{(0)} \rangle \quad (2)$$

and then aim to find an explicit expression for E_2 as a sum over terms, ideally of a similar form to $\sum_i -\frac{|(D_i|\hat{V}|\Psi^{(0)})|^2}{E_i - E_0}$ as appears in HCI+PT with $|D_i\rangle$ being a determinant, which can then be sampled stochastically. In our previous paper,²⁶ we chose \hat{H}_0 as

$$\hat{H}_0 = P E_0 P + Q \hat{H}_d Q, \quad (3)$$

where \hat{H}_d contains those operators in \hat{H} which do not change the occupation numbers of spatial orbitals, i.e.,

$$\hat{H}_d = \sum_p h_{pp} \hat{E}_{pp} + \frac{1}{2} \sum_{pq} (pp|qq) \hat{e}_{pqqp} + \frac{1}{2} \sum_{p \neq q} (pq|qp) \hat{e}_{pqpq}, \quad (4)$$

with $\hat{E}_{pq} = \sum_{\sigma} a_{p\sigma}^{\dagger} a_{q\sigma}$ and $\hat{e}_{pqrs} = \sum_{\sigma, \tau} a_{p\sigma}^{\dagger} a_{q\tau}^{\dagger} a_{r\tau} a_{s\sigma} = E_{ps} E_{qr} - \delta_{qs} E_{pr}$. This Hamiltonian is spin-free; however, it is only block-diagonal in the determinant space since it contains additional couplings for determinants with the same spatial occupations.²⁶ Thus, in this work, we instead used the Epstein-Nesbet (EN) form, which is equivalent to neglecting the above off-diagonal couplings in the determinant space. In the following, we will use \hat{H}_d to denote the EN form, which satisfies $\langle D_i | \hat{H}_d | D_j \rangle = \delta_{ij} \langle D_i | \hat{H} | D_j \rangle \triangleq \delta_{ij} E_i$. This choice will result in a slight difference in the computed E_2 compared with our previous p-DMRG, usually found to be less than $1mE_h$ (vide post), and the difference will gradually decrease as M_0 increases.

The EN partition is, of course, commonly used in SCI+PT.^{27–35} In this case, $Q[Q(\hat{H}_0 - E_0)Q]^{-1}Q$ in (2) can be simplified to $Q(\hat{H}_d - E_0)^{-1}Q$ for $Q = \sum_i |D_i\rangle \langle D_i|$ and $[Q, \hat{H}_d] = 0$, giving $E_2 = \sum_i -\frac{|(D_i|\hat{V}|\Psi^{(0)})|^2}{E_i - E_0}$ as a sum over determinants in the Q -space, as mentioned above. However, in our case, $Q(\hat{H}_d - E_0)Q$ and hence $Q[Q(\hat{H}_0 - E_0)Q]^{-1}Q$ are not diagonal in the determinant space due to the more complicated structure of Q arising from the fact that the zeroth order MPS potentially can have nonzero overlap with every determinant. To derive a similar expression for E_2 , we first rewrite

$$\begin{aligned} Q(\hat{H}_0 - E_0)Q &= Q(\hat{H}_d - E_0)Q \triangleq X - Y, \\ X &= (\hat{H}_d - E_0), \\ Y &= PX + XP - PXP. \end{aligned} \quad (5)$$

In general, the operator $X = \hat{H}_d - E_0$ can be assumed to be invertible by properly choosing E_0 , and this important issue will be discussed later. Then, using the relation $Q(X - Y)^{-1}Q = QX^{-1}(1 - YX^{-1})^{-1}Q = QX^{-1}Q + QX^{-1}YX^{-1}Q + \dots$ and directly evaluating each term, we can find that the series can be greatly simplified as P is one-dimensional such that $Q(X - Y)^{-1}Q$ becomes

$$Q(Q(\hat{H}_0 - E_0)Q)^{-1}Q = QX^{-1}Q - \frac{QX^{-1}PX^{-1}Q}{\langle \Psi^{(0)} | X^{-1} | \Psi^{(0)} \rangle}, \quad (6)$$

which can be simply verified by multiplying the right-hand side with $Q(\hat{H}_0 - E_0)Q$ to get the identity operator Q in the Q -space. The last term is the correction for the fact that $Q(\hat{H}_d - E_0)^{-1}P$ is in general nonzero, which clearly vanishes in the SCI+PT case as $[Q, \hat{H}_d] = 0$. The result is also easily derived by using Löwdin partitioning for $X = (\hat{H}_d - E_0)$ and blockwise inversion to fold in the self-energy contribution of the single P space Hamiltonian element on the Q space denominators. Substituting Eq. (6) into Eq. (2) for E_2 and invoking the resolution of identity $\sum_i |D_i\rangle \langle D_i| = 1$, each term in E_2 can finally be formulated as a sum over determinants, viz.,

$$E_2 \triangleq A + \frac{|B|^2}{C}, \quad (7)$$

$$A \triangleq -\langle \Psi^{(0)} | \hat{V} Q X^{-1} Q \hat{V} | \Psi^{(0)} \rangle = -\sum_i \frac{|\langle \Psi^{(0)} | \hat{V} Q | D_i \rangle|^2}{E_i - E_0}, \quad (8)$$

$$B \triangleq \langle \Psi^{(0)} | \hat{V} Q X^{-1} | \Psi^{(0)} \rangle = \sum_i \frac{\langle \Psi^{(0)} | \hat{V} Q | D_i \rangle \langle D_i | \Psi^{(0)} \rangle}{E_i - E_0}, \quad (9)$$

$$C \triangleq \langle \Psi^{(0)} | X^{-1} | \Psi^{(0)} \rangle = \sum_i \frac{|\langle \Psi^{(0)} | D_i \rangle|^2}{E_i - E_0}. \quad (10)$$

These formulae constitute the final working equations for stochastic p-DMRG. Note that unlike in the case of SCI+PT, here the summation over $|D_i\rangle$ goes over *all* the determinants.

In the practical evaluation of A , B , and C , instead of sampling $|D_i\rangle$ uniformly, we use importance sampling to improve the efficiency. For simplicity, we first discuss the evaluation of the term C (10). It can be evaluated as $C = \langle \frac{1}{E_i - E_0} \rangle_{P_i = |\langle \Psi^{(0)} | D_i \rangle|^2}$, where the subscript indicates that the average of $\frac{1}{E_i - E_0}$ is taken with respect to the population of the determinants generated with the probability $P_i = |\langle \Psi^{(0)} | D_i \rangle|^2$. To achieve this, we use an algorithm similar to that in Refs. 36 and 37. Specifically, suppose $|\Psi^{(0)}\rangle$ is in the right canonical form,

$$|\Psi^{(0)}\rangle = \sum_{n_1 n_2 \dots n_K} C_{\alpha_1}^{n_1} [1] R_{\alpha_{K-1}}^{n_K} [K] |n_1 n_2 \dots n_K\rangle, \quad (11)$$

where $C_{\alpha_1}^{n_1} [1]$ and $R_{\alpha_{K-1}}^{n_K} [K]$ are matrices and $R_{\alpha_{k-1} \alpha_k}^{n_k} [k]$ are rank-3 tensors satisfying the right canonical condition $\sum_{n_k \alpha_k} R_{\alpha_{k-1} \alpha_k}^{n_k} [k] R_{\alpha_k \alpha_{k+1}}^{n_{k+1}} [k+1] = \delta_{\alpha_{k-1} \alpha_{k+1}}$, then a determinant $|D_i\rangle = |n_1 n_2 \dots n_K\rangle$ can be sampled according to P_i by a single sweep from left to right as follows: The first occupation number $n_1 \in \{|vac\rangle, |1_\beta\rangle, |1_\alpha\rangle, |1_\alpha 1_\beta\rangle\}$ is generated according to $p_1(n_1) \triangleq \sum_{\alpha_1} |C_{\alpha_1}^{n_1}|^2$, which satisfies $\sum_{n_1} p_1 = 1$ as $|\Psi^{(0)}\rangle$ is normalized. Given n_1 , at the second site, the generation probability for n_2 is defined as $p_2(n_2|n_1) \triangleq \sum_{\alpha_2} |(C^{n_1} R^{n_2})_{\alpha_2}|^2 / N_2$. Importantly, the normalization constant N_2 can be found as $N_2 = \sum_{n_2} p_2(n_2|n_1) = p_1(n_1)$ due to the right canonical property. Repeating this procedure recursively, the generation probability for n_k given $n_1 \dots n_{k-1}$ can be defined as $p_k(n_k|n_1 \dots n_{k-1}) \triangleq \sum_{\alpha_k} |(C^{n_1} R^{n_2} \dots R^{n_k})_{\alpha_k}|^2 / N_k$ with $N_k = p_{k-1}(n_{k-1}|n_1 \dots n_{k-2})$. Thus, after the occupation of the last site is chosen, the total generation probability is $P(n_1 \dots n_K) = p_1(n_1)p_2(n_2|n_1) \dots p_K(n_K|n_1 \dots n_{K-1}) = |(C^{n_1} R^{n_2} \dots R^{n_K})|^2 = |\langle \Psi^{(0)} | D_i \rangle|^2 = P_i$, equal to our target distribution. For a spin-adapted DMRG implementation,¹⁵ the determinants can be generated similarly, but at

each step the Clebsch-Gordon coefficient $C_{S_{k-1}M_{k-1}S_kM_k}^{S_kM_k}$ needs to be incorporated to map each reduced tensor to the full one.

For term A (8), it would be possible to employ the same strategy, if $Q\hat{V}|\Psi^{(0)}\rangle$ can be expressed as an MPS faithfully. However, since converting $Q\hat{V}|\Psi^{(0)}\rangle$ into an MPS exactly would incur a bond dimension of $O(K^2M_0)$, this becomes prohibitive for large active spaces. Thus, we first use a bond dimension M_1 [that is small compared with $O(K^2M_0)$] to compress $Q\hat{V}|\Psi^{(0)}\rangle$ variationally as an MPS, i.e., $|\Phi\rangle \approx Q\hat{V}|\Psi^{(0)}\rangle$. This approximate state $|\Phi\rangle$ is then used to define a generation probability for $|D_i\rangle$, $P_i = |\langle\Phi|D_i\rangle|^2$, for importance sampling, such that

$$A = -\left\langle \frac{|\langle\Psi^{(0)}|\hat{V}Q|D_i\rangle|^2}{P_i(E_i - E_0)} \right\rangle_{P_i=|\langle\Phi|D_i\rangle|^2}. \quad (12)$$

In the case that $|\Phi\rangle$ is a good approximation to $Q\hat{V}|\Psi^{(0)}\rangle$, an approximate estimator for A is just

$$A \approx -\left\langle \frac{1}{E_i - E_0} \right\rangle_{P_i=|\langle\Phi|D_i\rangle|^2}, \quad (13)$$

which becomes similar to the expression for term C but with respect to a different distribution. This approximation becomes exact in the limiting case that $|\Phi\rangle = Q\hat{V}|\Psi^{(0)}\rangle$. It deserves to be noted that in Eq. (12), there is a subtlety. For the equality to hold, the set of determinants $\mathcal{S} \triangleq \{|D_i\rangle : \langle\Phi|D_i\rangle \neq 0\}$ interacting with $|\Psi^{(0)}\rangle$ must include the set that interacts with $\langle\Psi^{(0)}|\hat{V}Q|D_i\rangle \neq 0$, otherwise some contributions to E_2 will be missing. This is guaranteed for sufficiently large M_1 , and fortunately, we found that to converge to chemical accuracy ca. $1mE_h$, the required M_1 can be much smaller than $O(K^2M_0)$ (vide post). For the term B (9), we simply evaluate it as $B = \left\langle \frac{\langle\Psi^{(0)}|\hat{V}Q|D_i\rangle\langle D_i|\Psi^{(0)}\rangle}{P_i(E_i - E_0)} \right\rangle_{P_i=|\langle\Phi|D_i\rangle|^2}$ since it turns out to be less important in most cases due to its small size and the quadratic dependence of E_2 on it.

The above algorithm constitutes a stochastic algorithm for p-DMRG. Its computational cost depends on two parts:

the compression for $Q\hat{V}|\Psi^{(0)}\rangle$ on the right-hand side of the first order equation (1), which scales as $O(K^3M_1^2M_0)$ assuming $M_1 \gg M_0$, and the cost for the stochastic evaluation of E_2 . The latter is dominated by the number of samples N_s times the cost for evaluating the matrix elements $\langle\Psi^{(0)}|\hat{V}Q|D_i\rangle$. We found the following scheme to be efficient at least for the systems investigated in this work. By using the identity $\langle\Psi^{(0)}|\hat{V}Q|D_i\rangle = \langle\Psi^{(0)}|\hat{H}Q|D_i\rangle = \langle\Psi^{(0)}|(H - E_{DMRG}^{(0)})|D_i\rangle = \sum_j \langle\Psi^{(0)}|D_j\rangle\langle D_j|(H - E_{DMRG}^{(0)})|D_i\rangle$, the evaluation is converted into $O(K^2N^2)$ evaluations of overlaps between determinants and the zeroth order wavefunction $\langle\Psi^{(0)}|D_j\rangle$, each of which scales as $O(KM_0^2)$. Thus, the total cost for the stochastic evaluation is $O(N_sN^2K^3M_0^2)$, which is formally higher than $O(N_sK^3M_0^2)$ if the matrix element $\langle\Psi^{(0)}|(H - E_{DMRG}^{(0)})|D_i\rangle$ is directly computed by considering $|D_i\rangle$ as an MPS with bond dimension one. However, the efficiency of our choice in practice may stem from the fact that in computing $\langle\Psi^{(0)}|D_j\rangle$, the sparsity in the MPS tensors can be utilized such that the actual computational time is less than the direct evaluation of $\langle\Psi^{(0)}|(H - E_{DMRG}^{(0)})|D_i\rangle$. A detailed comparison of these two choices for large orbital spaces will be presented in future. The overall time of the stochastic step is usually found to be much smaller than the compression step. Thus, the present stochastic algorithm is more efficient than the previous p-DMRG algorithm, which required the iterative solution of the first order equation Eq. (1), which scales as $O(K^2M_1^3 + K^3M_1^2M_0)$, and with a much larger M_1 than required in the present algorithm.

We now demonstrate the accuracy and efficiency of the stochastic algorithm in comparison with the previous deterministic p-DMRG algorithm for two prototypical molecules: C_2 at its equilibrium bond length 1.24253 \AA ³⁸ in the cc-pVDZ basis set³⁹ and Cr_2 at its equilibrium bond length 1.68 \AA ⁴⁰ in the Ahlrichs' SV basis⁴¹ and the cc-pVDZ-DK basis.⁴² For C_2 and Cr_2 in the SV basis, all electrons were correlated, resulting in orbital spaces with 12 electrons in 28 orbitals, (12e, 28o)

TABLE I. Total energy and bond dissociation energy D_e of C_2 (in E_h) in the cc-pVDZ basis set (12 electrons in 28 orbitals) obtained by DMRG and stochastic p-DMRG ($N_s = 36\,000$) with various M_0 . Values in parentheses are the statistical uncertainties.

M_0	DMRG	p-DMRG		Stochastic p-DMRG ^a	
		$\lambda = 0$	$\lambda = \frac{1}{2}$	$\lambda = 0$	$\lambda = \frac{1}{2}$
Total energies of C ₂					
50	-75.708 599	-75.729 676	-75.731 740	-75.730 32(8)	-75.732 53(9)
100	-75.724 500	-75.731 540	-75.732 029	-75.731 73(3)	-75.732 27(3)
200	-75.729 502	-75.731 845	-75.731 990	-75.731 919(8)	-75.732 065(9)
400	-75.731 380	-75.731 939	-75.731 966	-75.731 947(3)	-75.731 976(3)
4000	-75.731 960				
Bond dissociation energies of C ₂					
50	0.184 82	0.205 87	0.207 93	0.206 51	0.208 72
100	0.200 69	0.207 73	0.208 22	0.207 92	0.208 46
200	0.205 69	0.208 04	0.208 18	0.208 11	0.208 26
400	0.207 57	0.208 13	0.208 16	0.208 14	0.208 17
4000	0.208 15				

^aStochastic p-DMRG does not converge to p-DMRG with increasing sample sizes due to the use of different \hat{H}_0 (see text).

TABLE II. Values of different components in stochastic p-DMRG with $\lambda = 0$ for C_2 (in E_h) in the cc-pVDZ basis set: E_2 , A , B , and C in Eqs. (7)–(10), respectively.

M_0	A	B	C	E_2
50	-0.021 82(8)	-0.000 9(1)	2.521(5)	-0.021 82(8)
100	-0.007 23(3)	0.001 5(1)	2.407(4)	-0.007 23(3)
200	-0.002 418(8)	0.000 03(5)	2.353(4)	-0.002 418(8)
400	-0.000 567(3)	-0.000 007(9)	2.343(4)	-0.000 567(3)

for C_2 and (48e,42o) for Cr_2 , respectively. For Cr_2 in the cc-pVDZ-DK basis, the same orbital space as in our previous work²⁶ was employed, viz., (28e, 76o) with the 1s, 2s, and 2p frozen. As in our previous work on deterministic p-DMRG,²⁶ the zeroth order energy is defined by an interpolation of two limits,

$$\begin{aligned} E_0(\lambda) &= (1 - \lambda)E_{DMRG}^{(0)} + \lambda E_d^{(0)}, \\ E_{DMRG}^{(0)} &= \langle \Psi^{(0)} | \hat{H} | \Psi^{(0)} \rangle, \\ E_d^{(0)} &= \langle \Psi^{(0)} | \hat{H}_d | \Psi^{(0)} \rangle. \end{aligned} \quad (14)$$

For simplicity, in this work, we only explored two cases, $\lambda = 0$ and $\lambda = \frac{1}{2}$, where in the former case $E_0(\lambda = 0) = E_{DMRG}^{(0)}$, and in the latter case $E_0(\lambda = \frac{1}{2})$ is an average $(E_{DMRG}^{(0)} + E_d^{(0)})/2$, which in practice we found to deliver significantly better energies for challenging systems due to the more balanced treatment of the zeroth order state and the determinant-like perturbors.²⁶ The limit $\lambda = 1$ was previously found to be prone to the intruder state problem and not considered here.

The results for C_2 obtained with stochastic p-DMRG with $N_s = 36\,000$ are listed in Table I together with the p-DMRG results. (Note that we report N_s as the total number of samples across all computational cores in the calculation, not the number of samples per core as is sometimes reported.) The orbitals were obtained by a complete active space self-consistent field

(CASSCF) calculation with an active space CAS(6e,6o). It is seen that both p-DMRG and its stochastic variant improve on the variational DMRG results at each M_0 , and converge toward the exact value computed by variational DMRG with $M = 4000$ very quickly. The $\lambda = \frac{1}{2}$ energies are better than those with $\lambda = 0$ due to the reasons discussed earlier. The bond dimension M_1 used for compression is chosen as 2000, which is sufficient for Eq. (12) to hold. The total energies of stochastic p-DMRG are seen to be very close to those from its deterministic counterpart for each λ . Neglecting off-diagonal couplings in Eq. (4), which is equivalent to employing the EN partition, in the stochastic variant, leads to slightly lower energies. This can be rationalized by the fact that after diagonalization of $Q\hat{H}_dQ$ these couplings will make the singlet perturbors have higher energies in the original p-DMRG compared with E_i for determinants, and hence the original p-DMRG will have smaller second-order energies for singlet states. It is seen that in all cases, the statistical errors are very small with a moderate number of samples. The wall time at $M_0 = 400$ is about 1 min for stochastic p-DMRG, which is 4 times faster than the deterministic one.

Next, the different components of E_2 , viz., A , B , and C in Eqs. (8)–(10), are given in Table II. The B term is found to be very small such that the $|B|^2/C$ contribution is negligible for E_2 in this case. However, since the calculation of B and C is relatively cheap compared with the evaluation of A , we always include them in all our calculations.

The Cr_2 dimer is a more difficult problem than C_2 and can be taken as a prototype of a challenging molecule in quantum chemistry. The corresponding results are shown in Table III and Fig. 1, with orbitals determined by a CASSCF calculation with CAS(12e,12o). In the stochastic calculations, the bond dimension for compression is $M_1 = 8000$, and the number of samples is $N_s = 28\,000$ except for one column where 10 times more samples were used to illustrate the convergence of the stochastic error. From Table III, we see that the

TABLE III. Total energy ($E + 2086$ in E_h) of Cr_2 in the Ahlrichs' SV basis (48 electrons in 42 orbitals) obtained by DMRG and stochastic p-DMRG.

DMRG energy (E in E_h) and discarded weights (w) with a reversed sweep schedule.						
M	500	1000	2000	4000	8000	∞
w	5.4×10^{-4}	2.1×10^{-4}	1.1×10^{-4}	5.8×10^{-5}	2.8×10^{-5}	
E	-0.4955	-0.5052	-0.5108	-0.5140	-0.5158	-0.5174(3) ^a
Stochastic p-DMRG energy ($N_s = 28\,000$) obtained with 28 000 samples. ^b						
M_0	DMRG	p-DMRG		Stochastic p-DMRG ^c		
		$\lambda = 0$	$\lambda = \frac{1}{2}$	$\lambda = 0$	$\lambda = \frac{1}{2}$	$\lambda = \frac{1}{2}$ ^d
100	-0.4236	-0.5000	-0.5147	-0.5025(3)	-0.5177(4)	-0.5182(1)
200	-0.4568	-0.5068	-0.5154	-0.5072(2)	-0.5167(3)	-0.5167(1)
300	-0.4785	-0.5107	-0.5161	-0.5118(2)	-0.5170(2)	-0.51726(8)
400	-0.4861	-0.5123	-0.5165	-0.5130(2)	-0.5171(2)	-0.51728(7)
500	-0.4921	-0.5136	-0.5169	-0.5143(2)	-0.5174(1)	-0.51751(6)

^aThe error bar for the extrapolated energy (∞) is estimated¹⁶ as 1/5 of the difference between the extrapolation energy and $M = 8000$ energy.

^bAll calculations were performed on a 28-core node. Values in parentheses are the statistical uncertainties.

^cStochastic p-DMRG does not converge to p-DMRG with increasing sample sizes due to the use of different \hat{H}_0 (see text).

^dResults obtained with 10 times more samples.

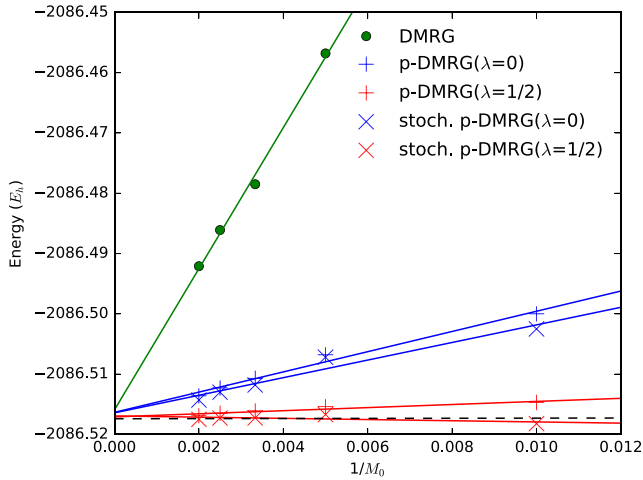


FIG. 1. Convergence of variational DMRG, p-DMRG, and stochastic p-DMRG energies as a function of M_0 for Cr_2 in Ahlrichs' SV basis. Both p-DMRG and stochastic p-DMRG improve the convergence significantly, and the choice $\lambda = \frac{1}{2}$ gives better results than $\lambda = 0$.

usual variational DMRG converges very slowly, and even with $M = 8000$ the error is about $1mE_h$ compared with the extrapolated result. By contrast, the convergence to chemical accuracy using p-DMRG and its stochastic variant is extremely fast using $\lambda = \frac{1}{2}$. As shown in Table III, for p-DMRG, chemical accuracy is achieved at $M_0 = 300$, while for stochastic p-DMRG it is achieved at $M_0 = 200$ (as its energy is usually slightly lower due to the same reasons as discussed for C_2). The convergence with $\lambda = 0$ is considerably slower compared with the $\lambda = \frac{1}{2}$ case, which we found in our previous study also, due to the unbalanced treatment of the zeroth order state and determinant-like perturbors.²⁶ However, it is still significantly cheaper than the variational DMRG for the same level of accuracy due to the smaller bond dimension M_0 . All stochastic evaluations took less than 15 min (wall time on a 28-core node) for a statistical error of less than $1mE_h$, while the variational compression for $Q\hat{V}|\Psi^{(0)}\rangle$ took about 30 min with $M_1 = 8000$. Thus, in total, the stochastic algorithm is about 5 times faster than the deterministic algorithm and further parallelizes much better than the deterministic one and uses much less memory and disk.

We found, in practice, that the bond dimension M_1 for compression does not need to be too large to achieve chemical accuracy. In Table IV, the dependence of the stochastic p-DMRG results on M_1 for $|\Phi\rangle$ are shown for $M_0 = 500$, $\lambda = 0$, and $N_s = 28\,000$. Using Eq. (12), $M_1 = 4000$ is sufficient to converge to $1mE_h$ accuracy. For too small M_1 ,

TABLE IV. Dependence of the total energy ($E + 2086$ in E_h) obtained by stochastic p-DMRG on the bond dimension M_1 used to compress $Q\hat{V}|\Psi^{(0)}\rangle$ as an MPS $|\Phi\rangle$. The results are illustrated with $M_0 = 500$ and $\lambda = 0$.

M_1	E_2 from Eq. (12)	E_2 from Eq. (13)
1000	-0.5100(15)	-0.50144(3)
2000	-0.5122(14)	-0.50735(6)
4000	-0.5143(7)	-0.51151(7)
6000	-0.5148(7)	-0.51272(8)
8000	-0.5143(2)	-0.51330(8)

TABLE V. Total energies ($E + 2099$ in E_h) of Cr_2 in the cc-pVDZ-DK basis set (28 electrons in 76 orbitals) computed by deterministic and stochastic p-DMRG with $\lambda = \frac{1}{2}$.

M_0	1000	2000	3000	4000
DMRG	-0.8346	-0.8617	-0.8743	-0.8818
p-DMRGa ($M_1 = 5 \times 7500$)	-0.9036	-0.9035	-0.9035	-0.9037
p-DMRGb ($M_1 = \infty$)	-0.9080	-0.9109	-0.9129	-0.9141
stochastic p-DMRG	-0.905(2)	-0.909(2)	-0.909(1)	-0.911(1)

^aReference 26: energy for $|\Psi^{(1)}\rangle = \sum_{i=1}^5 |\Psi_i^{(1)}\rangle$ with $M_1 = 7500$.

^bReference 26: extrapolated energy for $M_1 = \infty$.

due to the problem of a poor importance sampling function, which limits the set of the interacting determinants to be explored, the energy is slightly higher than the converged ones with larger M_1 . It is interesting to see that the approximate estimate (13) requires much larger bond dimension, approximately by a factor of 4, to achieve the same level of accuracy as Eq. (12), although its statistical error is smaller with the same N_s . Therefore, in practice, this approximation is not very useful.

Finally, we compare the performance of the stochastic p-DMRG against the deterministic p-DMRG for a larger problem, Cr_2 in an orbital space of (28e, 76o), in Table V. For this system, the convergence of the previous deterministic p-DMRG is very slow. Even with a sum of five MPS with bond dimension $M_1 = 7500$ for the first-order wavefunction, $|\Psi^{(1)}\rangle = \sum_{i=1}^5 |\Psi_i^{(1)}\rangle$, the deterministic p-DMRG energies differ from the extrapolated energy ($M_1 = \infty$) by 5–10 mE_h depending on M_0 . However, the energies provided by the stochastic p-DMRG with a given $M_1 = 10\,000$ and $N_s = 80\,000$ are already very close to the extrapolated $M_1 = \infty$ p-DMRG energies, demonstrating the efficiency of the stochastic p-DMRG. As we showed in our deterministic p-DMRG calculations,²⁶ one can further extrapolate in M_0 and M_1 to obtain an estimate of the exact energy without any truncation error. In future, we will explore the possibility of such extrapolations with stochastic p-DMRG also.

In summary, we have presented an efficient stochastic algorithm to overcome the previous bottleneck in p-DMRG,²⁶ developed recently for problems with large active spaces. We demonstrated that, in combination with a good choice of zeroth order Hamiltonian, the stochastic p-DMRG algorithm can provide highly accurate total energies for challenging systems with significantly reduced computational resources as compared with the deterministic p-DMRG, and both of these are much cheaper than the original variational DMRG, in large orbital spaces with a mix of correlation strengths.

Note: During the finalization of this work, we became aware of a related study in Ref. 43. This work assumed that E_2 can be approximated by the term A in Eq. (8) and used a different way to compute E_2 where $|\Psi^{(0)}\rangle$ was represented stochastically along the lines of the original stochastic HCI+PT.³⁴ The systematic accuracy of the final energies from that algorithm should be comparable to the p-DMRG energies with the choice $\lambda = 0$, although the statistical properties will not be the same.

ACKNOWLEDGMENTS

This work was supported by the US National Science Foundation through No. CHE 1665333. Additional support was provided by No. OAC 1657286. Z.L. is supported by the Simons Collaboration on the Many-Electron Problem. G.K.-L.C. is a Simons Investigator in Physics.

- ¹K. Andersson, P. A. Malmqvist, B. O. Roos, A. J. Sadlej, and K. Wolinski, *J. Phys. Chem.* **94**, 5483 (1990).
- ²B. O. Roos, K. Andersson, M. P. Flscher, P.-Å. Malmqvist, L. Serrano-Andrs, K. Pierloot, and M. Merchn, in *Advances in Chemical Physics*, edited by I. Prigogine and S. A. Rice (John Wiley & Sons, Inc., 1996), pp. 219–331.
- ³C. Angeli, R. Cimiraglia, S. Evangelisti, T. Leininger, and J.-P. Malrieu, *J. Chem. Phys.* **114**, 10252 (2001).
- ⁴C. Angeli, R. Cimiraglia, and J.-P. Malrieu, *Chem. Phys. Lett.* **350**, 297 (2001).
- ⁵C. Angeli, R. Cimiraglia, and J.-P. Malrieu, *J. Chem. Phys.* **117**, 9138 (2002).
- ⁶K. Andersson and B. O. Roos, *Chem. Phys. Lett.* **191**, 507 (1992).
- ⁷Y. Kurashige and T. Yanai, *J. Chem. Phys.* **135**, 094104 (2011).
- ⁸S. Guo, M. A. Watson, W. Hu, Q. Sun, and G. K.-L. Chan, *J. Chem. Theory Comput.* **12**, 1583 (2016).
- ⁹S. R. White, *Phys. Rev. Lett.* **69**, 2863 (1992).
- ¹⁰S. R. White, *Phys. Rev. B* **48**, 10345 (1993).
- ¹¹S. R. White and R. L. Martin, *J. Chem. Phys.* **110**, 4127 (1999).
- ¹²A. O. Mitrushenkov, G. Fano, F. Ortolani, R. Linguerri, and P. Palmieri, *J. Chem. Phys.* **115**, 6815 (2001).
- ¹³G. K.-L. Chan and M. Head-Gordon, *J. Chem. Phys.* **116**, 4462 (2002).
- ¹⁴Ö. Legeza, J. Röder, and B. A. Hess, *Phys. Rev. B* **67**, 125114 (2003).
- ¹⁵S. Sharma and G. K.-L. Chan, *J. Chem. Phys.* **136**, 124121 (2012).
- ¹⁶R. Olivares-Amaya, W. Hu, N. Nakatani, S. Sharma, J. Yang, and G. K.-L. Chan, *J. Chem. Phys.* **142**, 034102 (2015).
- ¹⁷S. Keller, M. Dolfi, M. Troyer, and M. Reiher, *J. Chem. Phys.* **143**, 244118 (2015).
- ¹⁸T. Yanai, Y. Kurashige, W. Mizukami, J. Chalupský, T. N. Lan, and M. Saitow, *Int. J. Quantum Chem.* **115**, 283 (2015).
- ¹⁹G. K.-L. Chan, A. Keselman, N. Nakatani, Z. Li, and S. R. White, *J. Chem. Phys.* **145**, 014102 (2016).
- ²⁰S. Sharma and G. K.-L. Chan, *J. Chem. Phys.* **141**, 111101 (2014).
- ²¹N. Nakatani and S. Guo, *J. Chem. Phys.* **146**, 094102 (2017).
- ²²S. Sharma, G. Knizia, S. Guo, and A. Alavi, *J. Chem. Theory Comput.* **13**, 488 (2017).
- ²³A. Y. Sokolov and G. K.-L. Chan, *J. Chem. Phys.* **144**, 064102 (2016).
- ²⁴L. Freitag, S. Knecht, C. Angeli, and M. Reiher, *J. Chem. Theory Comput.* **13**, 451 (2017).
- ²⁵A. Y. Sokolov, S. Guo, E. Ronca, and G. K.-L. Chan, *J. Chem. Phys.* **146**, 244102 (2017).
- ²⁶S. Guo, Z. Li, and G. K.-L. Chan, e-print [arXiv:1803.07150](https://arxiv.org/abs/1803.07150).
- ²⁷B. Huron, J. Malrieu, and P. Rancurel, *J. Chem. Phys.* **58**, 5745 (1973).
- ²⁸R. J. Buenker and S. D. Peyerimhoff, *Theor. Chem. Acc.* **35**, 33 (1974).
- ²⁹R. J. Harrison, *J. Chem. Phys.* **94**, 5021 (1991).
- ³⁰J. B. Schriber and F. A. Evangelista, *J. Chem. Phys.* **144**, 161106 (2016).
- ³¹N. M. Tubman, J. Lee, T. Y. Takeshita, M. Head-Gordon, and K. B. Whaley, *J. Chem. Phys.* **145**, 044112 (2016).
- ³²W. Liu and M. R. Hoffmann, *J. Chem. Theory Comput.* **12**, 1169 (2016).
- ³³A. A. Holmes, N. M. Tubman, and C. J. Umrigar, *J. Chem. Theory Comput.* **12**, 3674 (2016).
- ³⁴S. Sharma, A. A. Holmes, G. Jeanmairet, A. Alavi, and C. J. Umrigar, *J. Chem. Theory Comput.* **13**, 1595 (2017).
- ³⁵Y. Garniron, A. Scemama, P.-F. Loos, and M. Caffarel, *J. Chem. Phys.* **147**, 034101 (2017).
- ³⁶S. R. White, *Phys. Rev. Lett.* **102**, 190601 (2009).
- ³⁷E. M. Stoudenmire and S. R. White, *New J. Phys.* **12**, 055026 (2010).
- ³⁸M. Douay, R. Nietmann, and P. Bernath, *J. Mol. Spectrosc.* **131**, 250 (1988).
- ³⁹R. A. Kendall, T. H. Dunning, Jr., and R. J. Harrison, *J. Chem. Phys.* **96**, 6796 (1992).
- ⁴⁰V. E. Bondybey and J. H. English, *Chem. Phys. Lett.* **94**, 443 (1983).
- ⁴¹A. Schäfer, H. Horn, and R. Ahlrichs, *J. Chem. Phys.* **97**, 2571 (1992).
- ⁴²N. B. Balabanov and K. A. Peterson, *J. Chem. Phys.* **123**, 064107 (2005).
- ⁴³S. Sharma, preprint [arXiv:1803.04341](https://arxiv.org/abs/1803.04341) (2018).